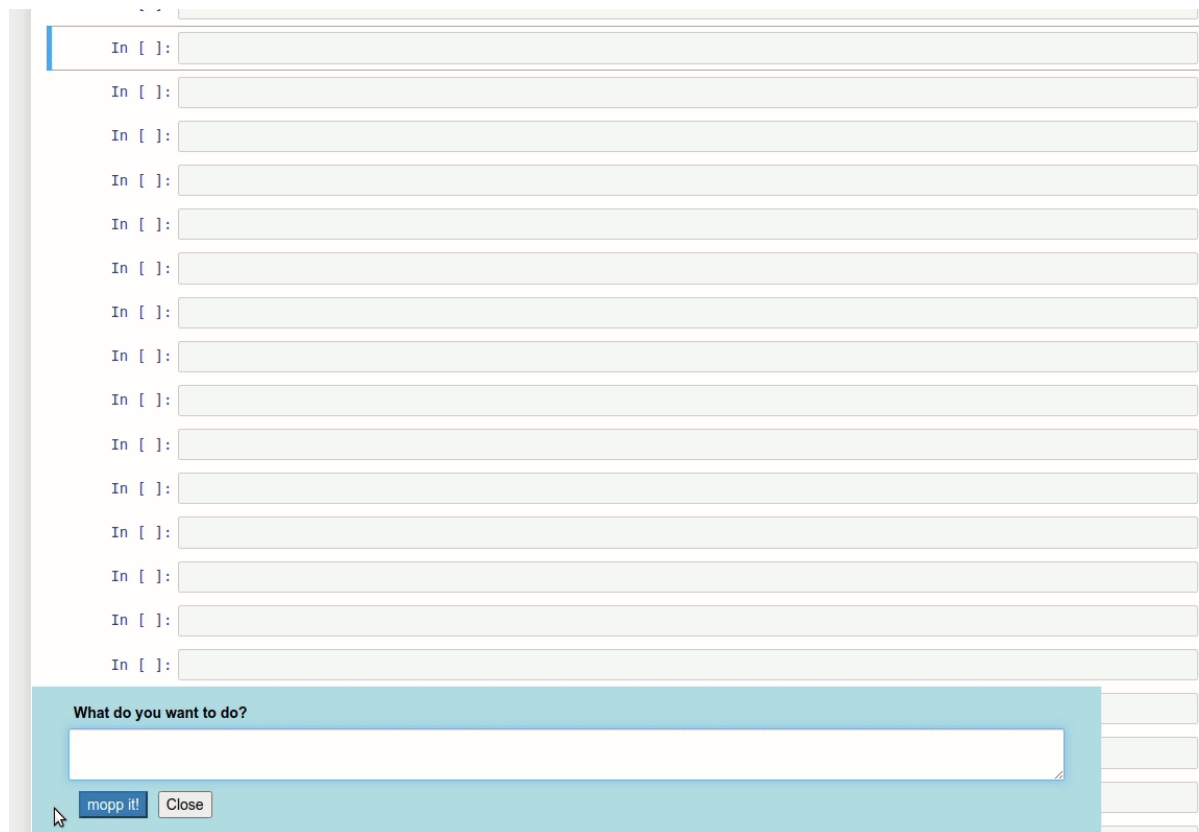


---

## Text2Code for Jupyter notebook

**A proof-of-concept jupyter extension which converts english queries into relevant python code.**



**Blog post with more details:**

**Data analysis made easy: Text2Code for Jupyter notebook**

**Demo Video:**

**Text2Code for Jupyter notebook**

**Supported Operating Systems:**

- Ubuntu
- macOS

---

## Installation

**NOTE:** We have renamed the plugin from mopp to jupyter-text2code. Uninstall mopp before installing new jupyter-text2code version.

```
1 pip uninstall mopp
```

**CPU-only install:** For Mac and other Ubuntu installations not having a nvidia GPU, we need to explicitly set an environment variable at time of install.

```
1 export JUPYTER_TEXT2CODE_MODE="cpu"
```

### GPU install dependencies:

```
1 sudo apt-get install libopenblas-dev libomp-dev
```

### Installation commands:

```
1 git clone https://github.com/deepklarity/jupyter-text2code.git
2 cd jupyter-text2code
3 pip install .
4 jupyter nbextension enable jupyter-text2code/main
```

## Uninstallation:

```
1 pip uninstall jupyter-text2code
```

## Usage Instructions:

- Start Jupyter notebook server by running the following command: `jupyter notebook`
- If you don't see `Nbextensions` tab in Jupyter notebook run the following command: `jupyter contrib nbextension install --user`
- You can open the sample `notebooks/ctds.ipynb` notebook for testing
- If installation happened successfully, then for the first time, Universal Sentence Encoder model will be downloaded from `tensorflow_hub`
- Click on the `Terminal` Icon which appears on the menu (to activate the extension)
- Type "help" to see a list of currently supported commands in the repo
- Watch Demo video for some examples

---

## Docker containers for jupyter-text2code (old version)

We have published CPU and GPU images to docker hub with all dependencies pre-installed. ##### Visit <https://hub.docker.com/r/deepklarity/jupyter-text2code/> to download the images and usage instructions.

**CPU image size: 1.51 GB**

**GPU image size: 2.56 GB**

## Model training:

The plugin now supports pandas commands + quick snippet insertion of available snippets from awesome-notebooks. With this change, we can now get snippets for most popular integrations from within the jupyter tab. eg: - Get followers count from twitter - Get stats about a story from instagram The detailed training steps are available in scripts README where we also evaluated performance of different models and ended up selecting SentenceTransformers [paraphrase-MiniLM-L6-v2](#)

## Steps to add more intents:

- Add more templates in [ner\\_templates](#) with a new intent\_id
- Generate training data. Modify [generate\\_training\\_data.py](#) if different generation techniques are needed or if introducing a new entity.
- Train intent index
- Train NER model
- modify [jupyter\\_text2code/jupyter\\_text2code\\_serverextension/\\_\\_init\\_\\_.py](#) with new intent's condition and add actual code for the intent
- Reinstall plugin by running: `pip install .`

## TODO:

- ☐ Add Ollama support to work with local LLMs
- ☒ Publish Docker image
- ☒ Refactor code and make it more modular, remove duplicate code, etc
- ☒ Add support for more commands
- ☒ Improve intent detection and NER

- 
- ☐ Add support for Windows
  - ☐ Explore sentence Paraphrasing to generate higher-quality training data
  - ☐ Gather real-world variable names, library names as opposed to randomly generating them
  - ☐ Try NER with a transformer-based model
  - ☐ With enough data, train a language model to directly do English->code like GPT-3 does, instead of having separate stages in the pipeline
  - ☐ Create a survey to collect linguistic data
  - ☐ Add Speech2Code support

**Authored By:**

- Deepak Rawat
- Kartik Godawat
- Abdullah Meda